

MemoWeb3



User manual

would like to thank you for getting the MemoWeb from the GOTO Software range. This does make you the owner of the Software but the liability of the right to use it under the conditions defined here after. By opening the sleeve containing the CD-ROM or by downloading the software, you accept, in reservation, and will conform to the terms and conditions of this License. If you do not accept the conditions in the contract, you must return the CD-ROM with its original packaging, as well as the accompanying material, to the place where you purchased the product, in order to receive a refund. Or you must return to download the product in the case of an on-line acquisition. The rights of the License holder in relation to software are stipulated in this License, excluding those who have not been expressly granted rights by GOTO.

TENTH OF THE LICENCE

With the exception of a contrary stipulation in an express GOTO, the software's user License is granted to a single computer in one place. Should you be in need, proof of purchase or acquisition (original CD-invoice or receipt) will be requested for each new installation of the Software. In the context of on-line acquisition, it is therefore advisable to print a corresponding invoice. The physical transfer of software from one computer to another is authorized and that the Software is only used on one computer. Only one backup copy is authorized and should be made out in accordance with article 2 below. The user is not authorized to copy the Software's printed content either, as intellectual property rights protect its use not stated in the previous paragraph is forbidden, as is any other use without written authorization from GOTO such as: Broadcasting, renting, selling, leasing, sub-licensing, giving, contributing all or part of the software to a third party, in any way, shape or form. Altering, disassembling, copying, adapting, creating derived works, integrating other software, translating or modifying in any way, in all or part of the Software. Carrying out of all or part of the Software (or anything other than the backup copy according to the terms and conditions defined here after) IN PARTICULAR IT IS FORBIDDEN THAT IT IS FORBIDDEN TO MAKE ALL OR ANY OF THE SOFTWARE AVAILABLE TO THE PUBLIC BY MEANS OF INTERNET, BBS OR MINITEL WITHOUT PRIOR WRITTEN AUTHORIZATION FROM GOTO.

II - BACKUP COPY

In the case of purchase of the CD-ROM used to operate the Software, this license authorizes you to have a backup copy in whatever form you choose. On the other hand, as soon as the program is copied onto the hard disk, the original CD-ROM will then be considered as the backup. In the case of on-line shopping, in addition to the software being installed on the hard disk, you are authorized to have a backup copy in whatever form you choose. The only purpose of this copy is to regenerate the original software in the event of accidental deterioration. On no account should it be used for any other purpose. Keep it in a safe place. Caution, your responsibility could be in question in the event of illicit use by a Third party. This license applies to the backup copy.

III - OWNERSHIP

GOTO owns all editing and reproduction rights for the Software and its documentation and on any authorized copy in the context of this license in accordance with the legislation on copyright royalties. Consequently you will take all necessary measures to protect said rights and in particular will respect and have respected all the essential information and warnings relative to these rights in relation to the elements making up the software or its backup copy.

IV - LENGTH - TERMINATION

This license is granted for an indefinite time except if you violate the terms and conditions of this license which will confer the option to GOTO to automatically cancel this license without prior notice and without prejudice for any other recourse it could undertake. In the event of cancellation, you are committed to ceasing all use of the software as well as its backup copy should there be one, and returning it to GOTO, without delay and upon GOTO's simple request.

V - GUARANTEE AND RESPONSIBILITY

GOTO in no way guarantees that the Software meets any particular functionality of any sort nor the user's specifications, be they in writing or not, prior to purchase. GOTO cannot be held responsible for damage or problems caused by the software under license and that could result from its use. This includes in a non-exhaustive way computing material, software, operating systems or any other computing accessory whatever it may be. The guarantee is strictly limited to support material that will be exchanged should a flaw be discovered. This guarantee which is valid for a period of 6 months from the date of purchase would be extended in the event of the product being returned, postage being chargeable to the sender.

NOTWITHSTANDING THE ABOVE, GOTO WILL NOT BE HELD RESPONSIBLE FOR ANY INJURY WHATSOEVER, DIRECT OR INDIRECT, CAUSED TO YOU OR TO A THIRD PARTY, IN PARTICULAR, AND PURELY AS GENERAL GUIDE AND WITHOUT LIMIT: LOSS OF REVENUE, EXPENSES, LOSS OF SOFTWARE, DETERIORATION OR LOSS OF SUPPORT MATERIAL BY YOURSELF OR A THIRD PARTY, INCLUDING RECLAMATION FEES, REPRODUCTION, OR COMPENSATION FOR SUCH LOSS OR DETERIORATION. You will indemnify GOTO for any damages resulting from GOTO being implicated and originating from a breach in this license or in the use of the Software.

VI - USER ASSISTANCE

This license does not oblige GOTO to assist in any way concerning the use of the Software. The user will find additional information, permanent help for the user and will be alerted to developments and new versions of the software either at the address <http://www.goto-software.com> or at the web site devoted to the product if there is one.

VII - PRECAUTIONS TO BE TAKEN WHEN USING A VIDEO GAME

It is recommended that people susceptible to physical disorder due to prolonged use of video games consult a doctor before any use. If disorder occurs during use (unconsciousness, dizziness, involuntary muscular contractions, blurred vision or disorientation, convulsions) immediately stop working and consult a doctor. In any event:

1. Check the room where your computer is has sufficient lighting
2. Do not sit too close to the screen especially if you have a large size screen
3. Position yourself so that you are looking down at the screen
4. Take a ten to fifteen minute break every hour

VIII - GENERALITIES

This license is subject to French law. The court dealing with any claim relative to the product's use or interpretation of this license is either the Paris court of first instance if the user is a private individual, or the Paris commercial court if the user is a company. GOTO will engage in proceedings against anyone infringing upon the terms of this license.

CONTENTS

Introducing MemoWeb 3	5
What is MemoWeb used for?	5
Who is MemoWeb for?	6
How does it Work?	7
What are the System Requirements for MemoWeb?	7
The Strong Points of MemoWeb 3	9
Capturing Web Sites	10
Fast Captures	10
Preparation	10
Capture Progress	13
Results	13
Advanced Captures	14
Limiting the Search of the Site's HTML Pages	14
Exploration Levels: How do you Explore Web Sites?	14
Downloading Document Types	16
Analyzing HTML Pages and Scripts	16
Other Options	17
Refining Capture Details	19
HTML Pages	19
Input Forms	20
Mapped Images	20
Inactive Links	21
Updating Captures.....	21
Setting Scheduled Captures	22
What is a Task?	22
Defining a New Task	22
Incorporating Captures to a Task	23
Modem Connection.....	23
Other Options on the Task Menu	23
Browsing the Contents of a Capture	24
"Search" Tool	24

Setting General Preferences	25
General	25
Simultaneous Connections	25
Limitations	25
HTTP and FTP Parameters	25
Internet Connection	26
Browser	26
Navigation Software	26
Shortcuts and Bookmarks	27
Proxy Server	27
Exclusion	28
The "Robot Exclusion" files	28
Exclusion List	28
Ad Banners Exclusion	28
Generation of Index Pages	28
Index of HTML Pages	28
Index of Images	28
Edit Reports	28
Options	29
Web Structure	29
Miscellaneous	29
Preview Window	29
Displaying Lists	29
Contacts	30
Glossary	31



INTRODUCING MEMOWEB 3

WHAT IS MEMOWEB USED FOR?

MemoWeb is a software with which you can recreate locally on your PC any Web site available on the Internet. MemoWeb connects itself for you, automatically surfs the Web, targets and stores on your hard disk everything that it encounters on the way: information pages, images, sounds, videos...

MemoWeb can download one or several Web sites while you are sleeping.

It retrieves the Web sites you are interested in at low cost during off-peak times: a time and money saver.

When it has completed its mission, you can take your time to browse offline the captured Web site in exactly the same way as if you were connected, with the best response times you could ever wish for.

Working in conjunction with your Web browser, MemoWeb multiplies the comfort of World Wide Web use.

MemoWeb mitigates the deficiencies of existing Web browsers: who has never tried painstakingly to rebuild locally a Web page with its images?

MemoWeb will do it for you in split seconds for a Web page, but also for a complete Web site and even for the whole of the Internet if your PC had enough memory!

MemoWeb is an **intelligent browsing agent** which surfs the Internet for you, according to the criteria you have defined. This agent has many qualities:

Fast:

MemoWeb reads the pages it meets and makes decisions on which actions to take faster than you do. It is also multitasking: it's as if you were to open simultaneously several navigator windows to the same Web site, without ever viewing the same page twice!

Obstinate:

Like all good explorers, MemoWeb doesn't stop half way but always goes to the end of the journey. Where you would lose patience when facing long response times on the net, MemoWeb persists.

Selective:

MemoWeb doesn't waste time consulting useless information.

It knows what you want and goes straight to it.

Methodical:

Nothing can escape it. Where you might not notice a link of interest, it cannot miss it. Conversely, it will never visit the same page twice.

When you are connected, how many times do you go back to the contents' page of a site?

Organized:

When its mission of exploration and capture has come to an end, MemoWeb generates additional index pages so that you can browse the Web site even more easily.

WHO IS MEMOWEB FOR?

For Everyone!

Because it is always more pleasant to browse a site offline, without having to wait at each page. Because MemoWeb allows you to browse the site anywhere at any time. Because MemoWeb optimizes connection times, especially when you don't have a permanent link with the Internet. Because MemoWeb never overlooks important information.

For Businesses with Local Networks

Following a single capture by MemoWeb all members of the network will be able to browse the Web site. In this way, a user can easily let others have access to his or her findings on the World Wide Web.

For the Teaching Profession

With MemoWeb, teachers can help their students discover the possibilities of the Internet in spite of restricted Internet connection facilities in school computer rooms.

For Web Site Developers

With MemoWeb, they can prepare offline, under the best possible conditions, demonstrations of their Web site designs for their clients and prospects.

The unpredictability of Internet connections is no longer part and parcel of the demonstrations.

For Web Site Administrators

With MemoWeb, they can detect very quickly any faulty link or missing file on the Web site. A single capture gives them an immediate and exhaustive portrayal of the state of a Web site.

For Image and Sound Collectors

For all those who wish to build up a collection of images or multimedia objects from the Internet, because MemoWeb is specially designed for the search, capture, organization and presentation of images available on the Internet.

For Lecturers

They can make use of HTML and Web sites to enhance their lectures.

For Multimedia Editors

MemoWeb can be used to easily convert an Internet site into CD-ROM multimedia, while still keeping part of the application online.

HOW DOES IT WORK?

To build a local Web site, you tell MemoWeb which are the entry points to that site, in the form of an HTML page Internet address (usually, this will be the home page of the Web site you want to capture).

MemoWeb connects itself to the Web just as you would do with your Web browser, sends the appropriate query string and downloads the HTML page you specified.

As soon as this page is completely downloaded, it analyses the contents of the page (HTML source) to detect two types of elements:

- Images or multimedia objects included in the page. Each of these objects is referenced in the page by its Internet address.
- Links pointing to other HTML pages, assigned to clickable areas on the screen. These links are analyzed according to the criteria provided to determine whether or not they should be explored.

These elements together lead to new query strings which MemoWeb sends to the Web server.

Each new HTML page downloaded is treated in the same way. The other files (images, sounds...) are simply stored on the hard disk. The Web site capture stops when there are no more pages to explore.

At the end of the capture, MemoWeb starts a procedure, known as Link Resolution, which entails reloading each HTML page captured and replacing the Internet address of each link and image by the filename of the corresponding local files.

In this way, MemoWeb rebuilds on your hard disk a completely independent Web site in which each HTML page points to other local pages.

The last phase consists in the creation of additional HTML pages which list the indexes to the various pages and images retrieved. These indexes allow an easier navigation through the local Web site.

Interactions with the Browser

MemoWeb is completely independent: it does not need the browser to capture pages, you can carry on surfing with your browser while MemoWeb is capturing sites. In fact, MemoWeb acts as a complement to your internet browser. Its **relationship** with the browser entails the following:

- Surfing a captured Web site: you can start surfing either from MemoWeb itself (button **Results**) or directly from the browser by loading the file **_Start.htm** in the Web capture folder.
- Choosing the addresses of Web sites to capture: at all times, you can ask MemoWeb to capture a site which you are currently browsing with your navigator. Indeed, when building a new Web site, by default MemoWeb will use the address of the site which is currently being visited by the browser.
- Creation of Internet shortcuts: each local Web site captured can be assigned a shortcut in your browser. This feature allows you to keep an updated list of all your local sites so that you can consult them more easily.

MemoWeb and the browser are interfaced using Dynamic Data Exchange (DDE) functions. The identification of the browser and its position on your hard disk are carried out automatically by default. However, you can disengage this mechanism and use several browsers (see Browsers p.26).

WHAT ARE THE SYSTEM REQUIREMENTS FOR MEMOWEB ?

- PC with Windows 95, 98, 2000 or Windows NT4
- Browser and Internet Connection
- 32 Mo of RAM
- 6 Mo available on hard disk (+ space for captured sites)
- CD-ROM Drive and Mouse

THE STRONG POINTS OF MEMOWEB 3

Grab any Website and surf it offline !

Thanks to MemoWeb 3, capturing Web sites is child's play. A few clicks of the mouse and you quickly download all that you are interested in onto your hard disk: texts, images, sounds and the fruit of the latest animation technologies on the Web, such as Shockwave, JavaScript, VRML or Java.

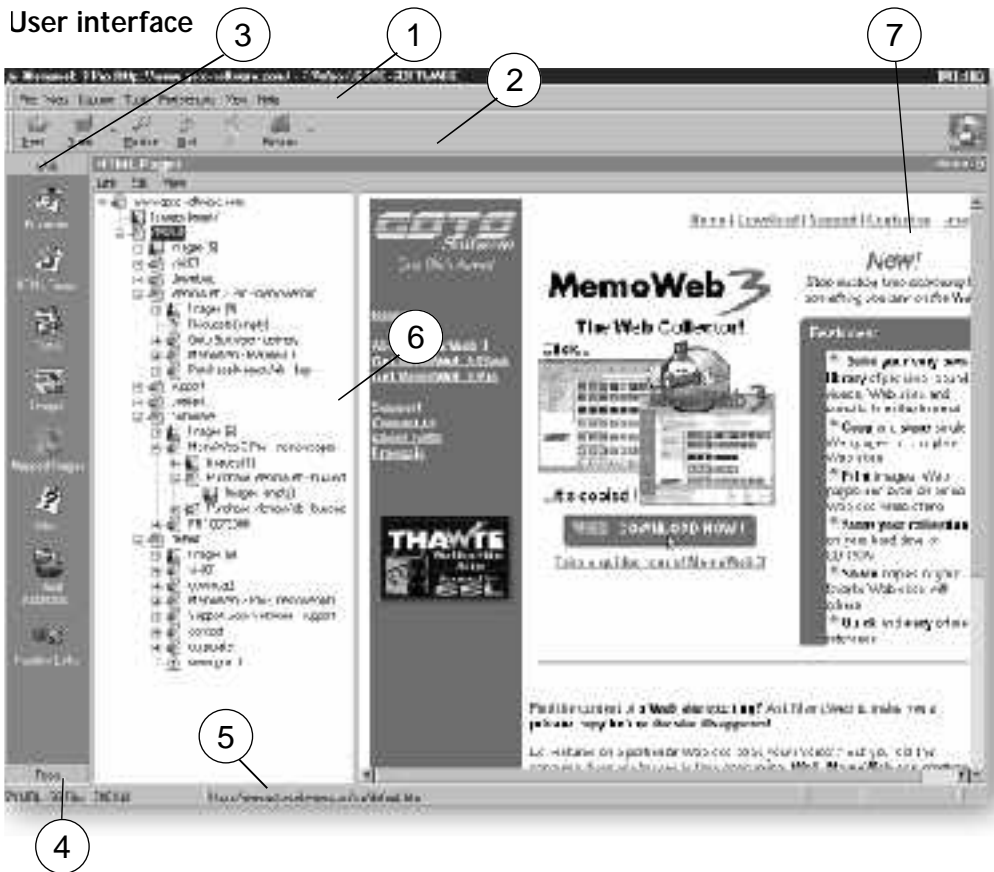
What's more, you benefit from the comfort of the latest features of MemoWeb 3:

- new user interface, with real time visualization of the hyperlink tree, and the possibility of immediate intervention during capture progress
- automatic and assisted processing of pages with input forms
- capture and export of e-mail addresses found
- optimized use of the transmission band for greater speed (up to 64 simultaneous downloaded files)

All of this in addition to features which were already part of previous MemoWeb versions:

- processing of mapped images
- scheduling capture times with possibility of running updates at regular intervals
- modem management to minimize Internet connection time.

User interface



1. The **menu bar** gathers most of the capture options, which are also available in the folders **Web** and **Tools**. For instance **File | New...** enables you to start a new capture.
2. The **tool bar** displays shortcut icons for the most useful options. To start capturing a site, press on **Grab**. To surf the captured site offline, press on the button **Results**.
3. The **folder Web** gathers the content of your captures (HTML pages, images, input forms, ...) and the capture settings, which can be accessed through the button **Properties**.
4. The **folder Tools** includes Search and Scheduled Tasks options, as well as a starting guide.
5. The **status bar** displays information about the captured site..
5. The **hyperlink tree** reflects the structure of the captured site.
7. The **preview window** displays the selected document in the hyperlink tree.

CAPTURING WEB SITES

Before starting to capture a Web site you must first define, as precisely as possible, the characteristics of the task you wish to entrust to MemoWeb.

This definition relies on 4 main criteria:

- Where do you want to start ?
- Where do you want to store data on your hard disk ?
- How and what do you want to capture ?
- When do you want to start capturing a site ?

You will see that each of these various criteria will be defined step by step when preparing for a capture.

For the time being, press on the **New** button or select **File | New...**

MemoWeb will now guide you through each step in the preparation of your capture.

FAST CAPTURES

Preparation

Where do you want to start from ?



This is the starting URL of your capture.

To start a capture, you must give MemoWeb an entry point in the Web network.

This entry point is usually a Web site home page, but can also be a page which is more deeply embedded in a server network.

The starting address of the capture is a URL (Uniform Resource Locator) comprising the names of a server, a folder and an HTML file.

Examples: <http://www.goto-software.com/>
 www.goto-software.com/us/index.htm (<http://> is optional)

If you want, you can give a name to your capture and then move on to the next step.



Tip: the pull down menu to the right of your browser's logo enables you to choose a starting URL from your Internet Explorer or Netscape Navigator bookmarks.

Where do you want to store data on your hard disk ?

MemoWeb saves the downloaded documents in a folder: this is the local Web site.

This folder contains:

- the retrieved files
- the file describing the structure of the site: MemoWeb also stores the directory structure of the captured site. It is this structure which allows it, amongst other things, to avoid trying to reopen the same file several times and following an interruption, to continue downloading a site from where it had left off. This file is always to be found in the capture folder, with the name of the **local.web**
- the contents' page (INDEX folder) and the general index page **_Start.htm**.

Select this destination folder with the button **Browse** and move on to the next step.

There are two ways of organizing a Web on the hard disk:

- flat: all the files of a Web are placed in the same folder which simplifies the copying of a site onto another support system (shared disk in a network, floppy, CD-ROM,...). In this case MemoWeb renames the files so as to avoid writing over files with the same name.
- in a website directory structure: MemoWeb recreates locally the folders and sub-folders which it meets on the servers as it explores the Web, adding a folder using the domain name (example: www.goto-software.com). To choose this last mode, check the box **Maintain the website directory structure** before starting a capture.

By default, MemoWeb gives extended names (Windows 95 format) to downloaded files unless the option **Use short filenames** is checked (when you start a capture). In this case, each retrieved file will have a name of 11 characters long in Windows 3.1 format (8 characters for the filename + 3 characters for the extension) made up from the original name used to specify this file on the Internet.

The complete structure of the captured Web site is saved in a file within the storage folder, the name of which is **local.web**. You must select this file with the option **Open** when you want to update a previously captured site .



How and What do you want to Capture ?

You are now going to define what you want to capture and how to do it.

- the **How** refers to the exploration limitations you want to impose on MemoWeb
- the **What** refers to the types of documents you want to capture.

To make the task easier for you, MemoWeb now offers standard settings which regroup both the **What** (images, sounds, HTML pages) and the **How** (folders, the complete site or the home page only).

For example, by choosing the setting **Images of the site**, MemoWeb will only download the images from the site defined by the URL entry point.

In fact, these standard settings are examples which you can customize at will. To define your own setting, change the capture settings with the button **Properties**, for example. Then, in the Web menu, select **List of standard settings** and click on **Save current setting**.



To start the capture, now click on **OK** and then the button **Grab**. MemoWeb then asks you to choose one of several options. Choose **Restart capture from scratch (deleting all local files if any)** to start the capture from the URL entry point.



Tip: by default, MemoWeb always suggests the option best suited to the type of capture you wish to run.

Capture Progress

This step is the least tiring for you !

Go and make yourself a cup of coffee or leave it in background mode while you work on something else with your PC.

MemoWeb will stop by itself when all the links have been explored, when all the documents have been retrieved or when a limitation defined for the capture has been reached.

When the capture is finished, MemoWeb carries out a process known as **Link Resolution**.

During this phase, it replaces the original links by links pointing to downloaded files or else pointing to an error page if the option **Redirect filtered links or links yet-to-be captured to a local page** has been checked.

During this process, the HTML index of your capture is generated.

Results

Press on the button **Results** to load the index in your browser. This will give you a precise account of the captured elements ordering the documents according to their type (HTML pages, Images, Files) and their status (Error). You can also load the summary page **_Start.htm** directly from your navigator.

Two other options are available:

- **Load captured site homepage**
- **Load bookmarks** (See Shortcuts and Bookmarks)



Tip: to load directly the home page of the captured site, select the option When pressing the button Results: Load capture index in the menu Preferences | Settings | Options.

If you wish to configure your capture in more detail, go to **Advanced Captures** now.

ADVANCED CAPTURES

We saw earlier (Fast Captures) how easy it is to prepare a capture using standard settings. If none of these answer your needs, you can define a new set, indicating the exploration limitations (search for HTML pages of the Web site) and the criteria for document downloading.

Limiting the Search of the Site's HTML Pages

What is Exploration ?

Let's begin at the URL address which you gave as the starting point of your capture.

MemoWeb first downloads the page found at this starting URL address and analyzes its contents to retrieve all other potential URLs (also known as **Link**) included in this page.

To know if, in turn, these links should be followed up, MemoWeb applies several filters which will either authorize or inhibit their exploration.

MemoWeb provides the possibility of applying a filter to link exploration; the links selected for exploration are defined by several criteria:

- link exploration levels
- document types to be retrieved
- analysis of HTML pages and scripts
- file type, quantity and size

Exploration Levels: How do you Explore Web Sites ?

Difference between Internal / External Links.

The initial WWW address is the exploration starting point.

It has an exploration internal depth level of 0.

A link is said to be internal if it belongs to the same Web site as the initial starting point address, i.e. if its URL belongs to the same domain as the base address (see the definition of URL).

The depth of navigation based on the number of internal link exploration levels

Each time an internal link is followed up, the internal exploration level is incremented.

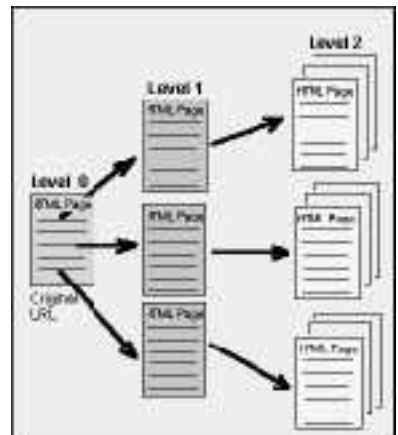
The pages linked to the initial page will therefore have a level 1.

The pages linked to a level 1 page will have a level 2, etc.

You can limit the exploration by setting the number of internal levels allowed in the **Exploration** document.

A level 0 is equivalent to the capture of a single page.

By default, the level is set at 255 which is in excess of the depth levels generally found in standard-sized Web sites.



The depth of navigation based on the number of external link exploration levels

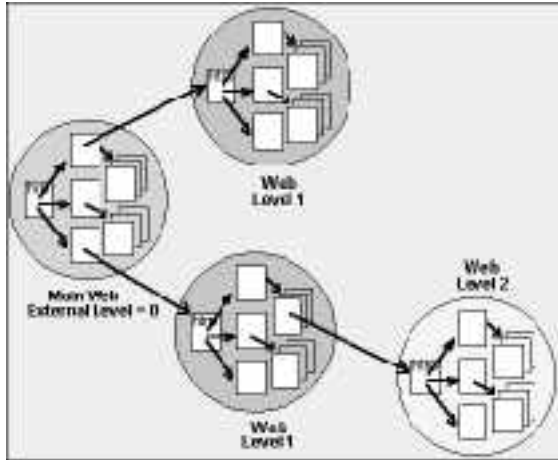
In much the same way as is the case for internal links, external links can be used to either restrict or extend your capture to other Web sites on the Internet.

A link is said to be external if it belongs to a different Web site to that with the initial starting point URL address, i.e. if its URL does not belong to the same domain as the base address.

Each time an external link is followed up, the external exploration level is incremented.

The sites linked to the original site will therefore have a level 1.

The sites linked to a level 1 site will have a level 2, etc.



The level of external exploration only concern HTML pages and do not affect other file types. If an HTML page includes an image belonging to another site, it will still be captured as if the external exploration level were 0.

This parameter can be set using the button **Properties**, tab **Exploration**.

By default, the level of external exploration is set at 0, which means that only the site with the initial starting point URL address will be captured.

For example:

MemoWeb retrieves the page with the starting point URL address <http://www.goto-software.com/us/index.htm>.

Amongst other things, it will find the following links on this page:

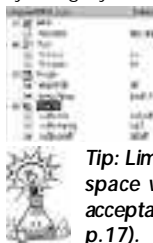
<http://www.goto-software.com/us/default.htm>

<http://www.goto-software.com/us/memoweb3/memoweb3.htm>

The two links refer to the same domain "goto-software", they will therefore have the same level of external exploration as that of the source page with the starting point URL, that is a level 0.

Downloading Document Types

When a remote server returns a requested file, it precedes it with a response header which contains information about the type of file, this is a **MIME type** declaration. The MIME type informs on the type of data enclosed in the file (this information cannot be derived from the file extension as it varies from one system to another). For example, HTML pages are of the type **text/html** and the image files in gif format are of the type **image/gif**. With MemoWeb you can define which file types you want to capture. Ten or so standard types are provided by default; you can either accept or inhibit the capture of any one of these. A file extension is assigned to each file type so that MemoWeb can name each captured file accordingly. The files are regrouped by category: HTML, text, images, sounds, video, scripts etc. but you can also define your own categories. To



do this, click or double click on a files category. In this way, you can also set a maximum or minimum file size for download. The box **Retrieve newly encountered types** is activated by default. This means that if MemoWeb comes across a file type that it doesn't recognize, it will save it and add this new type to the table of types. This table can be accessed after capture in the document **Web** | button **Properties** | tab **Files**.

Tip: Limiting document size. When retrieving a site, it is impossible to know in advance the disk space which will be required by the capture. Therefore, it is advisable to set a limit on acceptable document sizes or to restrict the number of files to be downloaded (see *Other Options* p.17).

Analyzing HTML Pages and Scripts

This step is subdivided into 3 parts:

- link resolution
- advanced parsing of Web documents (Java, Shockwave Flash and VRML)
- JavaScript code engine

Link Resolution

You can choose to **add a link (MemoWeb's logo) to the true end-of-page URL**. Take note that the logo can't be removed in MemoWeb 3 Flash version.

Or else you can **redirect filtered links or even yet-to-be captured links to a local page**. When MemoWeb finds a non-captured page, it replaces the link by a page which details the reasons for its absence, these could be:

- **_badlink.htm**: the link was not retrieved because the capture progress was interrupted
- **_error.htm**: any of the types of errors frequently encountered on the Internet (page not found, access denied...)
- **_filterXX.htm**: the link was filtered by exploration level limitations.

Finally, you can **generate automatically thumbnails for the captured pages**.

Advanced Parsing

Here you can define whether you would like MemoWeb to detect the required files and the existing links in Java, Shockwave Flash or VRML documents.

Warning! It is not possible to guarantee that these documents will work well locally as this is highly dependent on their internal structure.

Java: MemoWeb 3 will not only retrieve the JAVA applet (file with .CLASS extension) but also all other CLASS files required for this applet to run well. To obtain this, you need only ensure the box **JAVA applet** is checked.

VRML: 3D scenes are editable text files of the type VRML (Virtual Reality Modeling Language) with a .WRL extension.

MemoWeb analyses the contents of these scenes in the same way as it would an HTML page to extract the links to images, textures and other scenes associated to the document.

It then replaces the true URLs in this document by their equivalent addresses on the local web: this is the link resolution phase. You need only tick the box **VRML 1.0 and 2.0**.

Shockwave Flash: MemoWeb will retrieve the Flash animation .swf file and any other .swf file required for the animation to run correctly. You need only check the **Macromedia Shockwave Flash** box.

JavaScript Code Engine

Some HTML pages use the JavaScript language for elaborate events: such as changing the appearance of an image when the mouse moves over it (or an action implying onmouseover, onmouseout, onload), or else, for example, for input forms. There are 3 types of JavaScript:

- JavaScript which is triggered off automatically when an HTML page is loaded (for example, detecting the browser used to load pages with different content).
- Actions linked to events such as mouseover, clicking a link or loading a page. MemoWeb replaces the links in the documents analyzed by the local filename. In other words, it converts http:// files to local files.
- Simple input forms: MemoWeb can process simple input forms, i.e. those not requiring user input (scroll lists, check boxes, radio buttons). The other types of input forms can be processed after their initial capture (see Input Forms p.20).

In all cases, MemoWeb adapts the JavaScript code for offline execution.

Comments: since the various versions of JavaScript are not compatible with all navigators, a scroll menu offers the possibility of specifying which browser is to be used (Netscape or Internet Explorer).

Other Options

Here, you can define the options and limitations of your captures.

Capture Options

Specify here whether you would like to:

- retrieve background images
- only retrieve links pointing to other pages
- download FTP documents
- restrict the maximum duration of a capture
- avoid exploring pages containing frames

Limitations

To specify:

- The maximum number of files to be retrieved
- The maximum space to be occupied on the hard disk
- The minimum size of the files to be retrieved
- The maximum size of the files to be retrieved

We have reviewed all the parameters of the advanced settings for a capture. To start a capture, click on **OK** and ensure that you have checked the box **Start capture immediately** (bottom left).

If this box is not checked, you can start the capture later by clicking on the button **Grab**.

MemoWeb then provides the following options:

- Restart capture from scratch (deleting all local files if any)
- Complete a capture
- Update a completed capture
- Retry URLs in error and end the capture

You can also choose to disconnect automatically the modem at the end of the capture and to schedule a capture for a later date.

REFINING CAPTURE DETAILS

Following an initial capture, each document type retrieved by MemoWeb can be accessed independently by following the directory structure of the links, in much the same way as you would do with Windows "Explorer". You can access this feature through the menu **Web** and the buttons **HTML Pages**, **Forms**, **Images**, **Mapped Images**, **E-Mail Addresses** and **Inactive Links**.



HTML Pages

A directory structure presents the HTML pages encountered by MemoWeb during the capture session as well as their current status (page captured, yet-to-be captured or filtered by internal or external exploration level imitations).

MemoWeb offers ways of retrieving links according to the status of the documents concerned.

For example: facing the URL, the description indicates the status **Filtered by external Web exploration levels**. Click on this link ; in the popup window, MemoWeb will then suggest the following options:

- Force retrieval of the document
- Increase external domain exploration levels by 1
- Explore domain `www.domain_name.com`. You could for example ask for an exploration through a maximum of 7 levels.

Once you have made your choice, press **Apply** for your changes to be validated immediately.



The menu **Link** also allows you to specify whether or not you wish to capture some links by choosing to:

- force document retrieval (only applies to one link)
- inhibit document retrieval (only applies to one link)
- force capture of links included in the page (applies to all links).
- inhibit capture of links included in the page (applies to all links)
- neither force nor inhibit capture of links, cancel all operations tending to either force or inhibit.



Input Forms

Some pages can only be viewed once you have filled in and submitted a form. For example, the starting pages of search engines.

So that these can run locally, MemoWeb searches for all possible simple combinations.

You therefore only have to input the required data offline and submit it either by clicking on the send button provided in the form, or by right clicking and choosing **Submit query**. MemoWeb will then generate new URL addresses for capture.

If the capture has stopped, you will have to re-launch it. Locally, you will then need to fill in the input form to reach the relevant pages after downloading.

Note that a right click on the form also allows you to delete searches for pages which have not yet been captured, to reinitialize the input form and to generate queries automatically.



Mapped Images

The HTML display standard provides for a feature with which you can turn an image into a sort of multifunction button: an image can be divided up into geometric sections with each section assigned to a different link.

This capability is mainly used to display maps, however, it is not uncommon to find Web sites where hypertext choices have been replaced by images of this kind.

There are two types of clickable images:

- images with sections defined directly in the HTML source page containing the image (**client side**). When retrieving the HTML page, MemoWeb captures the definition of the sections and can therefore reproduce the mechanism in the local Web version.
- images with sections defined in a non-accessible file on the Web server (**server side**). In this case, when you click in the image, the browser sends the mouse coordinates to the server and this server acts on the coordinates to choose the appropriate link.

In the event of a capture, it would be impossible to simulate all possible positions of a mouse click in the image so as to rebuild the appropriate responses. MemoWeb uses a particular and exclusive technique to deal with clickable images of the server side type. One click on the button **Mapped Images** gives you the current list of clickable images included in the captured Web.

You can then load each of these images and define for yourself the sections which you think are hot points (see **Edit**).

Save these definitions when you have finished defining which are the hot points.

The third stage consists in starting a new capture. MemoWeb now retrieves the links assigned to each section and recreates locally clickable images of **client side** type.



Inactive Links

MemoWeb for one of several reasons may not have retrieved certain links, said to be inactive:

- the link was filtered by external Web exploration level limitations
- the link was filtered by internal Web exploration level limitations
- the page was not found (error 404)

See section HTML Pages above for more details.

UPDATING CAPTURES

After completing a first capture, when you press on the button **Grab**, MemoWeb will, depending on the status of your capture, suggest one of the following:

- Restart capture from scratch (deleting all local files if any)
- Complete a capture
- Update a completed capture
- Retry URLs in error and end the capture

Choose **Update the capture**

MemoWeb will start the process of Link reload.

For each file retrieved, it asks the server whether any changes have been made since the previous capture (by checking the date of the last modification to the file).

If there have been changes, it replaces the file by the new version and in the case of an HTML page, analyzes its contents to determine whether there are new links to explore.

Finally, when all the links have been reactualized, it carries out a process known as **Suppression of orphan links** which consists in the deletion of files which are no longer valid, thus avoiding unnecessary use of disk space.

Note ! This process will only start once all the links have been updated and there are no more files to retrieve. This is why, sometimes, you may have unused files stored temporarily in your capture folder.



SETTING SCHEDULED CAPTURES

One of the most interesting features of MemoWeb is the possibility of scheduling captures for a date and time of your choice without any further manipulation on your part.

To do this MemoWeb uses a scheduled capture administrator (when it is active, you can see its button on the Windows' applications taskbar).



The scheduled capture can be accessed from the menu **Tools | Scheduled tasks** or from the button **Scheduled tasks** on the tab **Tools**.

This scheduled capture administrator acts like MemoWeb's alarm clock:

- it will start MemoWeb if it is not already loaded in memory
- it handles **task** management

What is a Task ?

A task is a collection of events (capture, updating, etc.) which are run in sequence. There is only one task and one active event at any one time (because you cannot have more than one copy of MemoWeb loaded in memory at any one time).

The first process therefore entails Defining a New Task.

Defining a New Task

Select **Task | New task... | Start**.

- Task configuration is easy, and is summed up in 3 points:

Start When ?

Stop When ?

What do you want to do at the end of the task ?

Start When ?

You can choose to run a task monthly, weekly, daily or occasionally, all of these at any specified time.

So you can easily update a series of favorite Webs, on Monday mornings on arrival at the office or at noon during lunch breaks.

Stop When ?

You can define the maximum running time of the task (in minutes). This value is optional; if this box remains unchecked, the task will end when all its events have been finalized.

What do you wish to do at the end of the task ?

MemoWeb provides the option of cutting the modem connection and then closing itself.

Incorporating Captures to a Task

You only have to select **Task | New task...** | **Webs**.

A dialogue window appears, offering you the list of your last captures.

You can choose one or several (in add-mode to the Web list), and then select from the various types of task offered:

- Restart capture from scratch (deleting all local files if any): MemoWeb will start the capture from the beginning and will delete all existing files and links.
- Complete a capture: MemoWeb will retrieve those links which have not yet been captured.
- Update a completed capture : MemoWeb will update the links captured depending on their date values.
- Retry URLs in error and end the capture



***Tip:** Link resolution can take a long time, it can be preferable to do this offline to avoid incurring high connection rates. Link resolution can always be carried out later on. Finally, you can limit this task to a few minutes.*

Modem Connection

If you have checked the option **If the internet server connection is inactive...** in the folder **Preferences | Settings | Internet Connection**, then MemoWeb will automatically start the Internet connection at the beginning of the task; it will cut the connection again at the end if the box **Disconnect modem** is checked.

Other Options on the Task menu

Once the scheduled capture has been planned, you can interact with this schedule using the menu **Task**.





You can also:

- Delete the journal of scheduled tasks: after each scheduled capture MemoWeb keeps an ongoing journal of events.
- Delete, copy, and paste a task.
- Start, stop or suspend a task.



***Tip:** If you would like to minimize the application window during the execution of a scheduled capture, select **Minimize MemoWeb during capture** in the menu **Options** of the section **Scheduled tasks**.*

BROWSING THE CONTENTS OF A CAPTURE

Once a capture has been completed, you can explore the various types of files retrieved. From the menu **Web**, you can reach a listing of Images , E-mail  addresses, Inactive Links  and other types of Documents  (for example .mov video files) which have been captured.

All the documents can be viewed as a directory structure with links, in much the same way as when using Windows' Explorer. Note that with the star you can sort the updated documents.



Tip: *the selection of several HTML documents activates the option **Print Selection...** in the menu **Edit**.*

With respect to **Images**, **Inactive Links** and **other types of Documents**, a similar menu allows you to:

- open the document
- load the document in the browser
- find the parent pages linked to the document
- create a new Web from this URL, taking this page as the starting page of a new site
- save, copy, print the document
- create all the HTML thumbnails
- create a detailed error report on the inactive links

E-Mail addresses are immediately accessible from the menu **Address** from which you can:

- export the selected e-mails in text (.txt or .csv), HTML (.htm) or dBase III (.dbf) formats
- send a message to that address
- copy that address



"SEARCH" TOOL

VemoWeb 3 allows you to search for:

- parent pages of a link (pages containing this link)
- HTML pages according to defined textual content or title
- URLs using a filter

To do this, it uses a filter in text format which accepts the wildcards * and ?

* replaces any group of characters up to the one which follows the *

? replaces any unique character

You can very quickly find the parent pages of a link using the menus in the documents **HTML Pages**, **Images**, **Ignored Links**, etc.

SETTING GENERAL PREFERENCES

With the menu **Preferences | Settings...** you can configure MemoWeb's processing and display parameters.

GENERAL

Simultaneous Connections

- The maximum number of sockets corresponds to the number of connection links which can be opened in parallel. A socket corresponds to one line from the tracking zone in the Capture folder. The maximum number authorized is set at 16. If your machine is relatively slow, it is preferable to reduce the number of sockets to 3 or 4.
- The maximum timeout defines the delay following which a search request with no response is considered to have failed. This time corresponds to one request attempt, it is to be related to the number of connect attempts before error. By default, the maximum timeout is set at 30 seconds. In the event of high traffic on the network, frequently accessed servers or any other reason which could cause a decrease in the transmission band, it can be worthwhile to increase this delay.
- The number of connection attempts with an error refers to the number of reconnect attempts following an aborted search request.

Limitations

If this option is checked, MemoWeb will interrupt the capture when the number of errors encountered is greater than 30 (by default). In such cases, you can choose to start the capture again by clicking on the button **Grab Retry URLs in error and end the capture**.

Interrupting file download when the timeout between data transmission exceeds 30 seconds (by default): the server favors small files, it is therefore very likely that a file of several mega-octets would take a very long time to download. The best way to counter this slowness is therefore to define a limit to the timeout allowed between back downloads.

The average rate is defined along the same lines but concerns the average flow of file transfer.

Limiting the number of search requests per server to 3 (by default) allows a better use of the transmission band. When MemoWeb carries out a capture simultaneously on several servers, this capture is spread out more equally and therefore gains in efficiency.

Note: when operating on one server only, MemoWeb makes use of the maximum number of connections.

HTTP and FTP Parameters

Some Web servers require client software (agent) authentication in the form of a signature. Without this, they can reject the request (HTTP Request) for the document, leading to a capture error.

This authentication informs the server on the software it is in communication with so that it can provide an appropriate response.

Unless disallowed by your service provider, a scroll menu allows you to use the signature of a browser recognized by the servers (for example: Internet Explorer 5).

- If you have checked the box **Download FTP documents** (button **Properties | Options**), then MemoWeb will require an identifier when it connects to the FTP server (login). This identifier is characterized by a username and password. Usually, it carries out an FTP request of the anonymous public type (without password) but requires an e-mail address as login.

- The Option **Send requests using HTTP 1.1 protocol** theoretically improves the connection speed and hence retrieval of files for download. HTTP 1.1 protocol is supported by new generation servers and allows the capture to be carried out **in bursts**.
- With the option **Accept and use cookies**, MemoWeb will integrate systematically in all its search requests the cookies received from the servers. This feature is only useful with certain servers, to mimic the behavior of a browser.

Internet Connection

MemoWeb can establish automatically modem connection to your Internet access provider. To do this, it uses the remote access to the network provided by Windows. Choose within the list of available connections and check the box **Use the following connection**.



You can find the various modem access points in Windows with the menu **Start | Programs | Accessories | Remote Network Access**.

The connection is only effectively cut if you check the box **Disconnect modem** in the options offered when pressing on **Grab**, or if a scheduled task has come to an end.

Windows NT4 Characteristics:

With this operating system you must specify the path to the Windows telephone book (**Phone Book**). Usually, this file is called **RASPHONE.PBK** and is to be found in the **System32** folder.

Browser

Navigation Software

Browser detection

The navigation software (Web Browser) is normally detected automatically the first time MemoWeb is launched or a capture.

If the path to the browser does not appear in the typing zone, check the box **Automatically search for a registered browser** or use the button **Browse**.



When there are several browsers on the hard disk, the search will select automatically the one which was used last.

Finally, if you tick the box **Load browser at end of capture**, MemoWeb will load automatically the Contents **_Start.htm** into the browser after ending the **Link Resolution** phase.

Add a Grab! button in the Internet Explorer toolbar

You can customize the toolbar of your Internet browser with a **Grab!** shortcut launching MemoWeb.

This option is only available for Internet Explorer 5 (and further) browsers. It is unavailable if your default browser is not Internet Explorer.

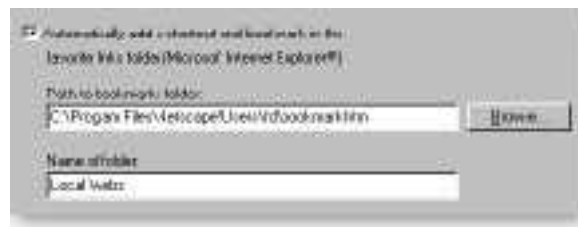
If the shortcut is not displayed in IE's Toolbar, you will have to add it manually by using the **Customize...** function located in the **View/Toolbars** menu.

Windows NT: In order to add a shortcut to Internet Explorer you have to be logged in as the superuser

Shortcuts and Bookmarks

The box **Automatically add a shortcut and bookmark** offers the possibility of integrating automatically:

- A bookmark for local Webs in the bookmarks' file of the browser, the bookmark will point to the home page of the captured site. In the case of Netscape Navigator, you only have to find the file **Bookmarks.htm** by using **Browse**.



- An Internet Link (file with the .URL extension) pointing to the home page in the **Favorites** Folder. Internet Explorer uses these favorite links.

By default, if Netscape Navigator is used, MemoWeb will check the existence of the file with bookmarks and of the folder for local webs. Should these not exist, it will create them automatically.

With this feature you can reach your local sites directly from the browser without having to launch MemoWeb.

Finally, you can define the name of the folder which will contain the capture shortcuts.

Proxy Server

A **Proxy** server acts like a local cache to the server. The Internet access providers use proxy servers for security reasons (if the access is obtained by a firewall) but mostly to increase the communication flow between client (yourself) and server, by storing the most consulted pages.

Depending on your Internet access provider, specifying a proxy can be compulsory, for example: Proxy address: proxy.your_provider.com Port:8080

Warning !

MemoWeb does not administer Proxy servers with an automatic configuration.

It is absolutely necessary to indicate a computer name (for example proxy.your_provider.com) or an IP address.

The proxy must only use the HTTP protocol, the other protocols (Socks) are not managed.

On an internal business network, your network administrator will be able to give you this information.

If your proxy uses a privileged access with username or password, specify these.

Characteristics of the Microsoft IIS Web Server

You must enter as identifiers the login and password input at the beginning of your NT network session.

If following several aborted attempts, the error message given by MemoWeb is of the type 407 (proxy authentication needed), contact your network administrator so that he can authorize the authentication known as **BASIC (cleartext)**. Indeed, by default the access to this type of server is allowed by way of owner encryption **NTLM authentication** of the login and password.

If the FTP proxy is the same (identical address and identical port number) as that of the HTTP, check the box **Use HTTP proxy for all FTP requests**.

Exclusion

The "Robot Exclusion" files

The exclusion file of a site limits the exploration of MemoWeb to the folders and documents authorized by the server which is questioned. The software configuration allows you to define whether or not you wish this to be taken into account or ignored.



Exclusion List

As you have already seen, MemoWeb uses the links between Web pages to restrict its exploration.

However, these exploration limitations are ineffective when:

- the exploration levels are high and you do not want MemoWeb to lose itself in a maze of links
- you only want to retrieve part of a Web, for example some folders on a given theme
- you want to avoid some sites (for example www.microsoft.com, www.yahoo.com, etc.)
- you want to retrieve a few files of a given type within a given folder (for example gif files : http://www.goto-software.com/*.gif).

Ad Banners Exclusion

This option (unchecked by default) enables you not to grab advertisement banners and to replace them locally with a transparent image.

Generation of Index Pages

Index of HTML Pages

MemoWeb can generate miniature reproductions of HTML pages: thumbnails. The format of these thumbnails ranges from 80x60 to 160x200 pixels. You can choose whether or not to keep the original background of the thumbnails (background image). In this case the default color is used. If you choose to display thumbnails while they are being generated, you can monitor the result but this will slow down the process. Since this is quite a lengthy process (as it entails loading the HTML page with its contents, resampling the image and compression in GIF format), it is advisable to run it once the site has been retrieved completely.

Index of Images

You can personalize the appearance of images in the index pages by changing the number of lines and columns as well as the size of the images.

Edit Reports

Modifies the number of lines of text for the other index pages: pages listing the various files and error pages.

Options

Web Structure

This will allow you to define the name of the root folder for all your captures. However, note that the choice of destination folder only comes into effect when you capture a new Web site. As soon as this is done, you can no longer change its position. By default, MemoWeb suggests a folder name which it builds using the default folder followed by a name derived from the Internet Web site address.

Examples: For the address www.goto-software.com, the name of the folder suggested will be Goto-software.
For the address www.goto-software.com/us/memoweb3, the name of the folder suggested will be MemoWeb3.

A folder on the hard disk can hold only one captured Web site.

If the folder which you specify does not yet exist on your disk, MemoWeb will create it automatically.

If it exists and already holds files from a captured Web site, MemoWeb offers to delete the existing files.

By default, the structure of the Web site (file local.Web) is updated on the disk during the capture process, once every 100 captured links.

Miscellaneous

Set MemoWeb to your tastes and habits:

- To the right of the button **Results**, a pull down menu will let you choose which page to display : the captured site home page, the capture index (report), or else the opened capture folder.
- Add a shortcut **Grab the website** in the menu **Start**: when you browse the Web, you only have to select **Grab the website** in the menu **Start** of Windows to initiate the immediate capture of the page you are currently browsing.
- Play sound at end of capture: define the sound to make.
- Define the background color of the folder menu-bars (gray by default).

Preview Window

Choose here whether to display the HTML pages in their standard format or as thumbnails in the MemoWeb preview window.

Displaying Lists

Choose the color of the links according to their current status, as well as the type and size of the font.

WWW.GOTO-SOFTWARE.COM/US/MEMOWEB3

Find MemoWeb on the Web !

THE LATEST UPDATES

www.goto-software.com/us/memoweb3/maj

FAQ (FREQUENTLY ASKED QUESTIONS)

www.goto-software.com/us/memoweb3/faq

GOTO Software reserves the right to change or delete without warning any information present on the site
www.goto-software.com

TECHNICAL QUESTIONS

By e-mail: hotline@memoweb.com

By fax: + 1 514 499 9202

Address: GOTO Software Inc. 410, rue Saint-Nicolas suite 09 - Montreal (Quebec) H2Y 2P5 - Canada

**FIND OUT ABOUT GOTO SOFTWARE PRODUCTS:
WWW.GOTO-SOFTWARE.COM**

Address:

see URL

CGI (Common Gateway Interface):

Programming language for the development of executable routines on Web servers responding to requests from HTML pages. The CGI routines are generally initiated by data which can be either specified by the HTML page itself, or input by the user (forms). CGI applications are countless and can range from deriving URL addresses from contextual parameters or displaying counters, to more complex database-related applications.

DNS (Domain Name Server):

Computer for an Internet site which processes requests by deriving the IP numbers from the requested URL addresses. Indeed, it is from these IP numbers that the routers (outside each link of the network) can transmit a request towards a given server.

Domain:

Information network connected on the Internet to which is assigned a name (domain name) referenced on the network as a whole. A domain name is associated to a class of IP numbers. The numbers of a given class have the same prefix and can identify each computer on the network.

FTP (File Transfer Protocol):

File transfer protocol over the Internet between an FTP server and an FTP client software.

HTML (Hyper Text Markup Language):

Description language of Web information pages. In particular it permits the inclusion of multimedia elements (images, sound, videos) and especially hyperlinks to other documents, whether or not they are on the same server. HTML pages are provided in text file format (source) by the Web server using the HTTP protocol and are decoded for display on the client screen by navigation software (Browser/Navigator).

HTTP (Hypertext Transfer Protocol):

Transfer protocol of data on the Internet between a Web server and a browser dedicated to HTML pages and their associated elements.

IP or TCP/IP (Internet Protocol):

Name of the transfer protocol of data between two remote networks which gave its name to the Internet network. It is a low level protocol which is complemented by a protocol specific to the type of application (HTTP, FTP, etc.).

SMAP:

Image with several links, the activation of which depends on the mouse cursor position at the time of the click. SMAP can be client side or server side. This name localizes the mapping between mouse coordinates and link URLs. When client side, this mapping is within the HTML page and the browser sends the request directly indicating the appropriate URL. When server side, this mapping is maintained on the server, the browser sends out a special request to the server indicating the mouse coordinates and the server returns the URL address to which it must connect.

JAVA:

Language designed for the development of applications (applets) sent to client computers as source code and executed on the client computer via the browser (which must therefore be JAVA enabled). The creators of the JAVA language (Sun) see as its main advantage the fact that it allows the development of applications independently of the operating system on the client computers.

Link:

Assigning a URL address to a clickable region on the screen. When the user clicks on this region, the page specified by the URL address is displayed. Since the page specified can be localized anywhere in the world, the act of going from one page to another by clicking on the links offered is known as surfing the Web. A link can be assigned indifferently to a sequence of characters (text which is usually underlined and changes color following activation) or to an image (see ISMAP for images with several links). In all cases, the browser detects the links in the pages and changes the appearance of the mouse cursor when it moves over them.

MIME (Multipurpose Internet Mail Extensions):

Classification of objects linked to HTML pages allowing their decoding by the navigation software. The various MIME types include sound, image and video files.

Navigator (or Browser):

Software designed to equip client computers on the Internet and allow them to send requests to Web servers and decode the HTML pages received. The better known navigators are Netscape Navigator and Microsoft Internet Explorer (MSIE).

IP number:

Identification number of a computer connected to the Internet network.

PROXY:

Computer on an Internet site acting as a cache for a given group of users. It stores on its own disk the information received from the network based on the users' requests. When it receives a new user request, it first checks for the presence of that information on its own disk and it is only in its absence that it will transmit the request to the network. A PROXY server significantly accelerates the consultation of the most requested servers.

Shockwave Flash:

Vectorial animation format used on the Internet and launched by Macromedia. It requires an appropriate plugin on your computer to be visualized.

Socket:

Logical communication between two systems linked to the Internet network. Contrary to analogical links (telephone) or to communications established physically between two users and for the duration of the communication, the numerical links are established only during data exchange between client and server.

A request sent from a client computer will open a socket with the server, which the server will close as soon as the required data has been sent. In the case of a modem connection, the closing of a socket is not synonymous with cutting the connection. A modem connection can allow the simultaneous opening of several sockets.

JRL (Uniform Resource Locator):

Syntax allowing the localization of a service on the network as well as the protocol used for dialogue with the application. For Web services which use the HTTP protocol, a URL address is of the type:

`http://computername.domainname.countryidentifier:portnumber/folder/filename`

The URL is the address of a service on the Internet.

VRML (Virtual Reality Modeling Language):

VRML is a system of navigation in three dimensions developed by Silicon Graphics and Intervista Software.

WEB or World Wide Web or again WWW:

The overall multimedia information services on the Internet using HTML presentation which include text, sound and fixed or animated images. By extension, Web can refer to any one of these services (the TF1 Web). Web server: host computer used to post Web services on the Internet.